Physica A 392 (2013) 3360–3373

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Benford's law: A Poisson perspective

Iddo I. Eliazar *

Holon Institute of Technology, P.O. Box 305, Holon 58102, Israel

HIGHLIGHTS

- Benford's law and its system-invariant generalization are explored via a Poisson-process setting.
- Poisson characterizations of Benford's law and its generalization are established.
- The universal emergence of Benford's law and its generalization is explained.
- A counterpart continued-fractions system-invariant distribution is studied.
- The power-law structure underlying Benford's law and its system-invariant generalization is unveiled.

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form 28 January 2013
Available online 11 April 2013

Keywords:
Power-laws
Invariance
Universality
Extreme-value statistics
Poisson limit theorems
Continued fractions

ABSTRACT

Benford's law is a counterintuitive statistical law asserting that the distribution of leading digits, taken from a large ensemble of positive numerical values that range over many orders of scale, is logarithmic rather than uniform (as intuition suggests). In this paper we explore Benford's law from a Poisson perspective, considering ensembles of positive numerical values governed by Poisson-process statistics. We show that this Poisson setting naturally accommodates Benford's law and: (i) establish a Poisson characterization and a Poisson multidigit-extension of Benford's law; (ii) study a system-invariant leading-digit distribution which generalizes Benford's law, and establish a Poisson characterization and a Poisson multidigit-extension of this distribution; (iii) explore the universal emergence of the system-invariant leading-digit distribution, couple this universal emergence to the universal emergence of the Weibull and Fréchet extreme-value distributions, and distinguish the special role of Benford's law in this universal emergence; (iv) study the continued-fractions counterpart of the system-invariant leading-digit distribution, and establish a Poisson characterization of this distribution; and (v) unveil the elemental connection between the system-invariant leading-digit distribution and its continued-fractions counterpart. This paper presents a panoramic Poisson approach to Benford's law, to its system-invariant generalization, and to its continued-fractions counterpart.

1. Introduction

Consider a database consisting of a large set of positive numerical values ranging over a span of orders, and observe the leading digit of each numerical value. For example, the leading digit of 731.26 is 7. What would you expect the statistics of the leading digits to be? Well, the common and intuitive answer is *uniform statistics*: we expect about 1/9 of the leading digits to be 1, about 1/9 of them to be 2, . . . , and about 1/9 of them to be 9. Alas, reality proves our intuition wrong. Indeed, in many real-world cases the statistics of the leading digits are markedly far from uniform.

The first published report on non-uniform statistics of leading digits was by the Canadian–American astronomer *Simon Newcomb* in 1881. Working with logarithmic tables, which were widely used in the pre-computer era, Newcomb noticed that the pages containing lower digits were more worn out than the pages containing higher digits. This phenomenon indicated

* Tel.: +972 507 290 650.
E-mail address: eliazar@post.tau.ac.il.

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.03.057

a deviation from uniformity which Newcomb investigated [1]. Newcomb's findings were re-discovered by the American physicist *Frank Benford* in 1938. Examining 20 different data sets – including 308 numbers contained in an issue of Readers' Digest, and the surface areas of 335 rivers – Benford convincingly exemplified and validated the prevalence of non-uniform statistics of leading digits [2].

Let us now quantitatively describe the non-uniform statistics discovered by Newcomb and Benford. To that end consider the representation of positive numerical values via a base β expansion, where β is an integer which is greater than one. For example, in the binary expansion the base is $\beta = 2$ and in the decimal expansion the base is $\beta = 10$. The base β expansion of a positive numerical value X is: $X = \sum_{r=-\infty}^{\infty} C_r \cdot \beta^r$, where $C_r \in \{0, 1, \ldots, \beta - 1\}$ is the expansion coefficient of rank r. The leading rank R of the numerical value X is the maximal rank for which the expansion coefficient is non-zero: $R = \max\{r \mid C_r > 0\}$. In turn, the leading digit D of the numerical value X is the expansion coefficient corresponding to the leading rank: $D = C_R$.

Consider now a large set of positive numerical values, and pick from the set a numerical value X at random. The randomization renders X and its leading digit D *random variables*. As noted above, intuition suggests that the distribution of the leading digit D is uniform over the integers $\{1, \ldots, \beta - 1\}$. However, Newcomb and Benford asserted that the distribution of the leading digit D is given by

$$\Pr(D = d) = \frac{\ln(d+1) - \ln(d)}{\ln(\beta)} = \log_\beta\left(1 + \frac{1}{d}\right) \tag{1}$$

($d = 1, \ldots, \beta - 1$). The leading-digit distribution of Eq. (1) is nowadays commonly termed *Benford's law*, and its cumulative distribution function is given by

$$\Pr(D < k) = \frac{\ln(k)}{\ln(\beta)} = \log_\beta(k) \tag{2}$$

($k = 1, \ldots, \beta$). Three excellent non-technical expositions of Benford's law include Refs. [3–5].

Benford's law is widely encountered in the sciences [6–9]. Specific examples of Benford's law in the physical sciences include complex atomic spectra [10], statistical physics [11,12], and seismic clusters [13]. Interestingly, Benford's law is successfully applied in forensic accounting and fraud detection [14–16]. The omnipresence of Benford's law has called out for theoretical understanding of mechanisms giving rise to this law. Theoretical explanations of Benford's law fall into three main categories: statistically implicit, statistically explicit, and physical.

The statistically-implicit explanations characterize Benford's law as the unique leading-digit distribution satisfying scale-invariance [17] and base-invariance [18]. Scale-invariance means that if we change the scale of the measurements – say by shifting from the English system to the metric system – then the leading-digit distribution will remain unchanged. Base-invariance means that if we change the base of the expansions – say by shifting from the decimal expansion ($\beta = 10$) to a hexadecimal expansion ($\beta = 16$) – then the logarithmic structure of Eqs. (1) and (2) will be maintained. The statistically-explicit explanations establish Benford's law as the outcome of probabilistic limit laws involving mixtures of distributions [19,20]. Informally, these limit laws assert that if we collect data from many different sources – each source producing data governed by a different probability distribution – then, for large samples, the data's leading-digit distribution will follow Benford's law. The physical explanations generate Benford's law via physical mechanisms such as multiplicative processes [21], exponential growth [21,22], and one-dimensional discrete dynamical systems [23]. Contrary to the statistically-implicit explanations, both the statistically-explicit explanations and the physical explanations provide constructive methods for the generation of Benford's law. For an interesting debate regarding whether or not there is a "simple explanation" to Benford's law the readers are referred to Refs. [24,25].

A key feature rendering the explanation of Benford's law non-trivial is the following intrinsic contradiction: on one hand, Benford's law implies that the distribution of the random variable $\ln(X)$ must be uniform; on the other hand, the random variable $\ln(X)$ takes values on the real line, and there is no uniform probability distribution over the real line. One way to overcome this intrinsic contradiction is to apply mixtures of many different distributions – as done in the statistically-explicit explanations [19,20] – thus, in effect, approximating a uniform probability distribution over the real line. A radically different way to overcome the intrinsic contradiction is to abolish altogether the underlying *probabilistic framework* and to shift to an alternative modeling framework that does accommodate uniform distributions over the real line: *Poisson processes*.

In this paper we follow the 'radical path' and study Benford's law from a *Poisson perspective*. Poisson processes are perhaps the most common statistical methodology to model the random scattering of points in general domains [26–28], and have applications ranging from queueing systems [29] to insurance and finance [30]. A key feature of Poisson processes is their ability to accommodate infinitely many orders of scale – a capacity which is well applicable in the context of: (i) statistical power-laws [31,32]; (ii) linear, nonlinear, and randomized central-limit-theorems for aggregates and extrema [33,34]; (iii) fractal stochastic processes [35,36]; and (iv) anomalous-diffusion phenomena [37]. In particular, this key feature of Poisson processes allows for the accommodation of uniform scattering over the real line – which appears to be an essential ingredient required for the modeling and generation of Benford's law.

In this paper we show that Poisson processes indeed offer a natural harbor to Benford's law. After a terse review of Poisson processes (Section 2) we:

- Establish a Poisson characterization of Benford's law (Section 3), and present a Poisson 'multidigit extension' of Benford's law (Section 4).